**SCHOLARSHIP STATISTICS & MODELLING: SOLUTIONS for Challenge Set 2.**

These problems involve a mixture of the work done so far this year in the externally assessed standards. There are seven problems this week.

**Problem 1 (from 2008 Scholarship Examination)**

**Bookings Analysis**

1. Binomial: n=100, ** = 0.3. P(*X* > 25) = 1 – P(*X* < 24)

**=** 1 – 0.11357

**= 0.88643.**

NOTE: Alternatively (if you don’t have a G.C.), use a **normal approximation** since Binomial tables don’t go high enough on *n*. Can approximate to normal since n** =30 which is > 5 *and* n(1-**)=70 which is >5.

Need to make a continuity correction, since approximating discrete with continuous. We get:

P(Bin: *X* > 25, *n*=100, *p*=0.3) ≈ P(Normal: *X* > 24.5, with **=*n**, *)

≈ 0.8849 (via normal dist tables).

1. Let *X*: Number of bookings that result from *n* phone calls.

P(10th phone call results in 5th booking)

= P(Binomial: 4 bookings from the 1st 9 phone calls) × P(10th phone call results in a booking)

=P(Binomial: *X* = 4, *n*=9, ** =***a*** ) × ***a***, since *a* is the probability that 10th one results in booking.

= 9C4 ***a***4(1-***a***)5 × ***a***

= 9C4 ***a***5(1-***a***)5 Note that this is an example of the *Negative Binomial Distribution*.

1. P(Booking) = P(Phone  booking) + P(Internet  booking) + (Fax  booking)

= 0.6***a*** + 0.3***a***2 + 0.1***a***3

And substituting in ***a*** = 0.7, we get:

P(Booking) = 0.6(0.7) + 0.3(0.7)2 + 0.1(0.7)3

= 0.6013

Hence P(No booking) = 1 – 0.6013

= 0.3987. ****39.87% of enquiries result in no booking.

1. In terms of ***a***, the probability that a randomly chosen enquiry results in a booking is:

P(Booking) = 0.6***a*** + 0.3***a***2 + 0.1***a***3.

P(Phone │ Booking) = 

= 

=  **NOTE:** In this situation, we ***can*** cancel out the common

factor of ***a***, without cutting out any possible solutions. This is because:

1. The only possible solution for ***a***  eliminated by

doing so would be that of ***a***=0, but we’re told that 0 < ***a*** <1. So ***a*** cannot be 0 anyway.

**2.** We can ignore the possibility of the

denominator being equal to 0, because we know

that this probability is *defined* (i.e. as 75%).

We’re told that P(Phone │ Booking) = 75%. So  ‘







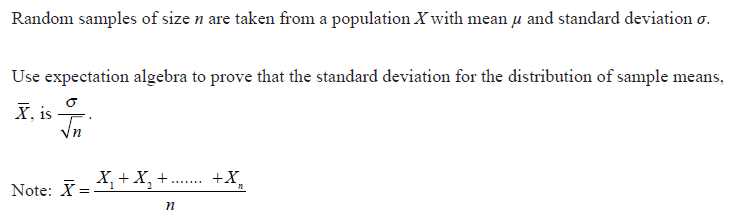




***a***  = 0.5616 (4sf), or -3.561.

**Given that 0 < *a*** < **1, then** ***a*  = 0.5616 (4sf)**

**Problem 2 – The Standard Error of the Sample Mean (from 2007 NCEA AS90642 – Excellence)**

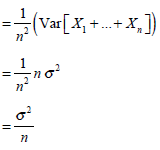


**Solution:**



*= *

Using the fact that Var(a*X*) = a2.Var(*X*)





And, since , hence as required.

**Problem 3 – distribution of sample means**

1. Mean of distribution of sample means for samples of n=36:  = ****(by the Central Limit Theorem)

= **250mL**

Standard deviation (standard error) of distribution of sample means for samples of n=36:

 = **** 

= **** 

= **mL or mL**

1. Calculate the probability that a random sample of 36 bottles would have a mean volume of 248.2mL or less if Speedy Sports Waters claims about the mean and standard deviation of the volume of drink in its bottles are true.

P( < 248.2 if =250 and **=5) = 



= 

= 0.5 - 

= 0.5 – 0.4846 (using tables)

= **0.0154** (tables) or 0.015386 (using G.C.). **i.e. only about 1.5%.**

1. What conclusion(s) do you draw based on your answer to (b)? What recommendation would you make to the manufacturers of Speedy Sports Water?

***Answer:* Only about 1.5% of random samples of size 36 would be expected to produce mean volumes as low as 248.2mL *if* in fact the *true* mean volume of all Speedy’s bottles is 250mL as claimed.**

**Hence I conclude that either I got a very unrepresentative sample or, what is much more likely, that the mean is not 250mL but something slightly lower.**

Another possibility to consider is whether the population standard deviation is much greater than the 5mL claimed. In other words, do the volumes vary from bottle to bottle more than what is claimed? However a few calculations tell us that the standard deviation would have to be more than 2 ½ times larger (over 12.5mL) for the probability of getting a sample with a mean this low to rise as high as just 20%. This is unlikely because commercially produced bottles are filled and sealed on production lines by machine, and this process means volume per bottle is unlikely to vary much.

**Hence we conclude that, most likely, the mean volume of drink per bottle is less than the claimed 250mL. This is dishonest marketing by Speedy. I recommend that the manufacturers increase the amount of drink that is dispensed into their bottles to match the claims of a mean volume of 250mL.**

**Problem 4 - exploring the standard error of the sample mean.**

1. Standard error of sample mean, **** *=* ****

Rearranging this to make  the subject:   ****

 ×3.18

= **19.08**

1. A normally distributed random variable has a standard deviation 1.44.

What sample size would be required to obtain samples whose means varied with a standard error of 0.72?

Standard error of sample mean, **** *=* ****

Rearranging this to make *n* the subject:  *=* ****

*****n =* ****

Substituting in ****=1.44 and =0.72 : *n =* ****

*=* 22

*=* 4. **sample size of 4 required.**

1. By what number is the standard error of the sample mean multiplied if a sample size is:
   1. Increased from 10 to 50?

**Bigger sample size means less variability in sample means, so  gets smaller.**

**** *=* **.** Soif *n* is multiplied by 5 (10 up to 50), **** must be divided by **.**

**which means multiplying it by** ** i.e. ×0.4472 (4sf).**

* 1. Decreased from 100 to 25?

**Smaller sample size means greater variability in sample means, so  gets larger.**

*****=* **.** Soif *n* is divided by 4 (100 down to 25), **** must be multiplied by **.**

**which means multiplying it by , i.e. ×2.**

1. A normally distributed population has mean 219 and standard deviation 18. If many random samples of size 100 are drawn from this population, predict the **upper quartile** of the distribution of sample means.

P(**<Upper Quartile) = 0.75 (by definition of an *upper quartile*).

**Hence** P(<**<Upper Quartile) = 0.25

P(0 < *Z* < z) = 0.25 gives a *z* value of 0.67448.



**So** *=* 0.674 (tables) (or 0.67445 on GC)



*=* 0.674



*=* 0.674



*=* 0.674



= 1.2132



= **220.2** (4sf) using tables. Same answer to 4sf using G.C.

**Problems 5🡪7 are about describing and critically evaluating the 4 main sampling techniques.**

**They do not come with solutions. They are to be handed in and marked by your teacher.**